EX. B

Gene, 111 (1992) 229–233
 0378-1119/92/$05.00

GENE 06296

Primary structure of the *Aequorea victoria* green-fluorescent protein

(Bioluminescence; Cnidaria; aequorin; energy transfer; chromophore; cloning)

Douglas C. Prasher[a], Virginia K. Eckenrode[b], William W. Ward[c], Frank G. Prendergast[d] and Milton J. Cormier[b]

[a] *Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543 (U.S.A.);* [b] *Biochemistry Department, University of Georgia, Athens, GA 30602 (U.S.A.) Tel. (404)542-1747;* [c] *Department of Biochemistry and Microbiology, Cook College, Rutgers University, New Brunswick, NJ 08903 (U.S.A.) Tel. (908)932-9562; and* [d] *Department of Biochemistry and Molecular Biology, Mayo Foundation, Rochester, MN 55905 (U.S.A.) Tel. (507)284-2065*

Received by S.R. Kushner: 21 March 1991
Revised/Accepted: 13 September/27 September 1991
Received at publishers: 26 November 1991

SUMMARY

Many cnidarians utilize green-fluorescent proteins (GFPs) as energy-transfer acceptors in bioluminescence. GFPs fluoresce in vivo upon receiving energy from either a luciferase-oxyluciferin excited-state complex or a Ca^{2+}-activated photoprotein. These highly fluorescent proteins are unique due to the chemical nature of their chromophore, which is comprised of modified amino acid (aa) residues within the polypeptide. This report describes the cloning and sequencing of both cDNA and genomic clones of GFP from the cnidarian, *Aequorea victoria*. The *gfp10* cDNA encodes a 238-aa-residue polypeptide with a calculated M_r of 26888. Comparison of *A. victoria* GFP genomic clones shows three different restriction enzyme patterns which suggests that at least three different genes are present in the *A. victoria* population at Friday Harbor, Washington. The *gfp* gene encoded by the λGFP2 genomic clone is comprised of at least three exons spread over 2.6 kb. The nucleotide sequences of the cDNA and the gene will aid in the elucidation of structure-function relationships in this unique class of proteins.

INTRODUCTION

Luminescence is common in a variety of marine invertebrates. Many cnidarians and probably all ctenophores emit light when mechanically disturbed. Proteins responsible for bioluminescence from several species of these two phyla have been characterized. Light from luminescent cnidaria is primarily green whereas light emitted from ctenophores is blue. The green light of cnidaria is due to the presence of a class of proteins called green-fluorescent proteins (GFPs). They are highly fluorescent and are activated in vivo by an energy transfer process via a luciferase or a Ca^{2+}-activated photoprotein, both of which produce energy during the oxidation of coelenterate-type luciferin. In the cnidarian *Aequorea*, the photoprotein aequorin excites the GFP by an unknown mechanism to release green light. Previous studies suggesting that *Aequorea* GFP is stimulated via a radiationless mechanism (Morise et al., 1974) have been questioned (Ward, 1979). The GFP from *Renilla*, another cnidarian, on the other hand, clearly receives energy from the *Renilla* luciferase-oxyluciferin excited state

Correspondence to: Dr. D.C. Prasher, Redfield Bldg., Woods Hole Oceanographic Institution, Woods Hole, MA 02543 (U.S.A.)
Tel. (508)457-2000, ext. 2311; Fax (508)457-2195.

Abbreviations: *A.*, *Aequorea*; aa, amino acid(s); bp, base pair(s); GFP, green-fluorescent protein; *gfp*, DNA or RNA encoding GFP; kb, kilobase(s) or 1000 bp; nt, nucleotide(s); oligo, oligodeoxyribonucleotide; ORF, open reading frame(s).

complex by a radiationless energy transfer mechanism (Ward and Cormier, 1976).

The GFPs most thoroughly studied have been isolated from *Aequorea* and *Renilla* (Ward, 1979). The *Aequorea* GFP has been reported to be a 30-kDa monomer (Prendergast and Mann, 1978) whereas the *Renilla* GFP is a 54-kDa homodimer (Ward and Cormier, 1979). The two proteins have different absorption spectra but identical emission spectra (λ_{max} = 509 nm). Upon denaturation the two GFPs have the same absorption spectra. Ward et al. (1980) have predicted that both *Aequorea* and *Renilla* GFPs contain chromophores having the same structure but that the different absorption spectra are explained by different apoprotein environments.

Biochemical properties of the *Aequorea* GFP show it to have unique structural properties. The fluorescent chromophore is stable to a variety of harsh conditions including heat, extreme pH, and chemical denaturants. Fluorescence is lost, for example, to base or acid treatment or addition of guanidine hydrochloride, but upon neutralization of the pH or removal of the denaturant, fluorescence returns with an identical emission spectrum (Bokman and Ward, 1981; Ward and Bokman, 1982). The chromophore structure is very different from those of the phycobiliproteins which are also highly fluorescent. The chromophore in the GFPs is covalently bound and is formed by modification of certain aa residues within the polypeptide. The chemical structure of the *Aequorea* GFP chromophore (Fig. 1), first characterized by Shimomura (1979), has been thoroughly re-examined (Ward et al., 1989; W.W.W., unpublished) and is shown here (Fig. 1) in its revised form. In this study, the *Aequorea* GFP gene and its cDNA have been isolated and characterized in pursuit of elucidating the mechanism of energy transfer between aequorin and GFP as well as addressing evolutionary relationships in coelenterate bioluminescence.

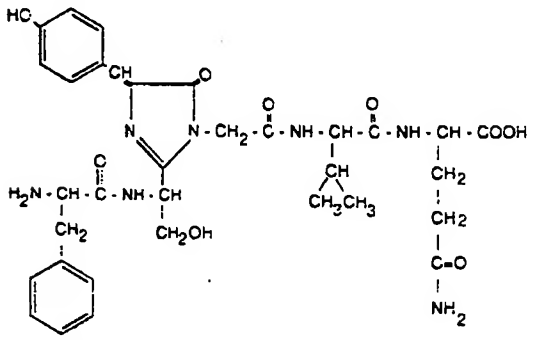

Fig. 1. The chemical structure of the chromophore in *Aequorea* GFP (W.W.W., unpublished). The cyclized chromophore is formed from the tripeptide Ser-dehydroTyr-Gly within the polypeptide by an unknown mechanism.

EXPERIMENTAL AND DISCUSSION

(a) Construction of cDNA libraries

An *A. victoria* cDNA library, constructed in pBR322 (Prasher et al., 1985), was screened for the presence of *gfp* cDNA using two oligo mixtures whose sequences were based on the aa sequences derived from GFP-derived CNBr fragments. The oligos contained the following nt sequences: **A**: 5′–AA$^{A}_{G}$AA$^{A}_{G}$TC$^{A}_{T}$TG$^{C}_{T}$TG$^{C}_{T}$TTCAT (20-mer with 32 redundancies). **B**: 5′–TT$^{A}_{G}$TA$^{A}_{G}$TT$^{A}_{G}$TA$^{C}_{T}$TCCAT (17-mer with 16 redundancies). The hybridization of the ^{32}P-labeled mixtures **A** and **B** to replicate filters containing this library were performed according to the method of Wood et al. (1985) utilizing tetramethylammonium chloride during the washing steps. The temperatures used during the washing steps for mixtures **A** and **B** were 55°C and 50°C, respectively.

A single *gfp* cDNA was isolated from the library by this method. This clone, pGFP1, contained a *Pst*I insert of 511 bp having an ORF encoding 168 aa. The deduced translation of the nt sequence indicated the *gfp1* cDNA lacked both the 5′- and 3′-sequences of the coding region. However, the sequence FSYGVQ within the deduced translation permitted the chromophore structure to be deciphered (W.W.W., unpublished). Upon rescreening the library with *gfp1* cDNA, no additional cDNAs were found.

A second *A. victoria* cDNA library was constructed (Gubler and Hoffman, 1983) in λgt10 (Huynh et al., 1985). The *Pst*I insert from *gfp1* cDNA was used as a hybridization probe against the entire λgt10 library of 1.4×10^6 recombinant phage. No *gfp*-related recombinants were identified upon screening the primary library. The phage remaining on the plates were extracted from the top agar and used as an amplified library (Maniatis et al., 1982). Upon screening this preparation of the library, four recombinants hybridized to the *gfp1* cDNA following their purification. The four cDNA clones were designated λGFP10, 11, 12, and 13. All four recombinants were shown to contain an insert of 1 kb upon digestion with *Eco*RI.

(b) Characterization of the *gfp10* cDNA

The entire *Eco*RI insert of λGFP10 was sequenced (Fig. 2). Limited nt sequences obtained from λGFP11 and 12 were identical with that from λGFP10 suggesting that they were siblings and, hence, were not sequenced further. Even though the entire coding region appears to be present (see below), three features of the cDNA insert of λGFP10 suggest it is not quite full-length. First, the cDNA is 0.5 nt where the *gfp* mRNA is 1.05 kb in length as determined by Northern analysis (Fig. 3). Second, the 5′–untranslated region is very short. Third, no poly(A) track is observed in the *gfp10* cDNA sequence (Fig. 2) despite the presence of the *gfp* mRNA in only the poly(A)$^+$ RNA fraction of *A. vic*

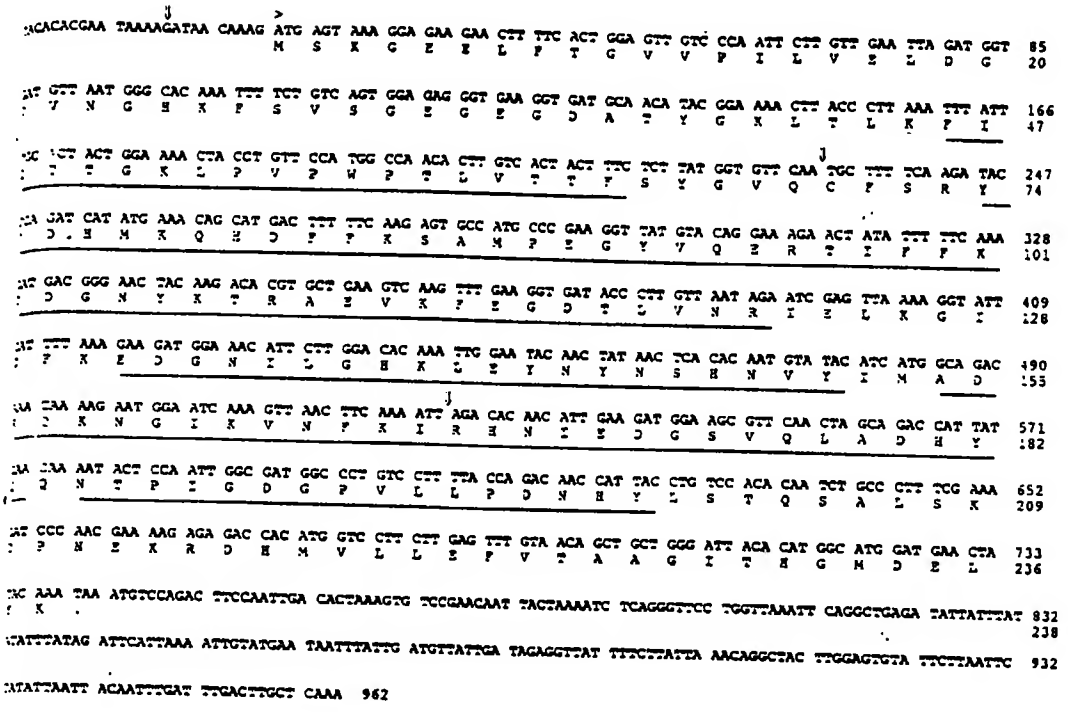

Fig. 2.

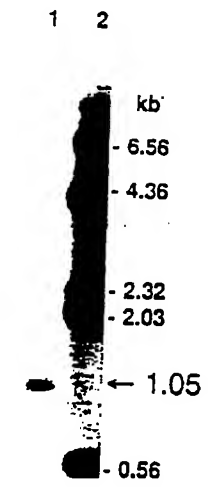

Fig. 3.

Fig. 2. Nucleotide sequence of the *gfp10* cDNA and the deduced aa sequence. Below the first nt of each codon is the single-letter designation for the aa. The horizontal lines underline those aa sequenced directly from native GFP. The downward arrows indicate the positions of introns when compared to the nt sequence of the *gfp2* gene. Arrowhead: start codon; period: stop codon. DNA fragments from both cDNA and genomic clones were subcloned into M13mp18 and M13mp19 (Yanisch-Perron et al., 1985), and unidirectional deletions were prepared using the method of Dale et al. (1985). Sequencing was performed using either the Klenow fragment or an altered T7 DNA polymerase (Sequenase Ver 2.0, United States Biochemical Corp.) in the dideoxy chain termination method (Sanger et al., 1977). Both DNA strands of the sequences described in this report here have been sequenced. The GenBank accession No. for the *gfp10* sequence is M62653.

Fig. 3. Northern analysis of the *A. victoria gfp* mRNA. The poly(A)$^+$ mRNA (lane 1) was denatured using glyoxal prior to electrophoresis, as described by Thomas (1983). Electrophoresis was performed for 3 h in a 1% agarose gel (pretreated with 10 mM sodium iodoacetate) equilibrated in 10 mM sodium phosphate pH 7.0 buffer. Overnight transfer of the nucleic acids to nitrocellulose was facilitated with 20 × SSC. Hybridization of ^{32}P-labeled *gfp1* cDNA to the membrane-bound nucleic acids was at 42°C for 28 h in 5 × SSC/5 × Denhardt's/20 mM Na-phosphate pH 6.8/100 μg per ml of denatured herring sperm DNA/10% polyethyleneglycol/50% formamide. *Hin*dIII-digested λ DNA, ^{32}P-labeled, and treated in parallel with the RNA, was used as molecular weight standards (lane 2).

toria RNA (data not shown). A typical polyadenylation signal is located at nt 861–865 (Fig. 2).

The nt sequence of the *gfp10* cDNA contains an ORF encoding a 238-aa protein having a calculated M_r of 26 888. This compares favorably with 30 kDa for native GFP as determined by denaturing electrophoresis (Prendergast and Mann, 1978). The deduced translation contains aa sequences of numerous peptides isolated from native GFP (underlined in Fig. 2). When compared to the *gfp10* cDNA sequence (Fig. 2), the *gfp1* cDNA was determined to encode aa residues 28–195. Oligo mixture A is complimentary to the codons encoding aa 78–84 and mixture B is complimentary to the codons encoding aa 141–146 (Fig. 2). The trimer Ser-Tyr-Gly, modified in the native protein to form the chromophore (W.W.W., unpublished), is located at aa 65–67. The chromophore consists of an imidazolone ring formed by the residues Ser-dehydroTyr-Gly within the polypeptide (Fig. 1). Located 8 aa upstream of this chromopeptide is GFP's only Trp. The inability to detect the fluorescence from this Trp makes it unusual (W.W.W., unpublished). Perhaps energy-transfer occurs between it and the chromophore in the native protein preventing the Trp fluorescence (320–350 nm). The Trp is flanked by several Pro residues (Pro-Val-Pro-Trp-Pro). The significance of this pentapeptide is not understood but a search of the protein databases (PIR ver 25; Swiss-Prot ver 14) shows it to be present only in cytochrome P-450 proteins.

(c) Isolation and characterization of *gfp* genomic clones

The *gfp1* cDNA was also used to isolate genomic clones prior to the availability of the *gfp10* cDNA. An *A. victoria* genomic library was constructed in λ2001 (Karn et al., 1984) essentially as described (Maniatis et al., 1982). Eight recombinant phages hybridizing to the *gfp1* cDNA were

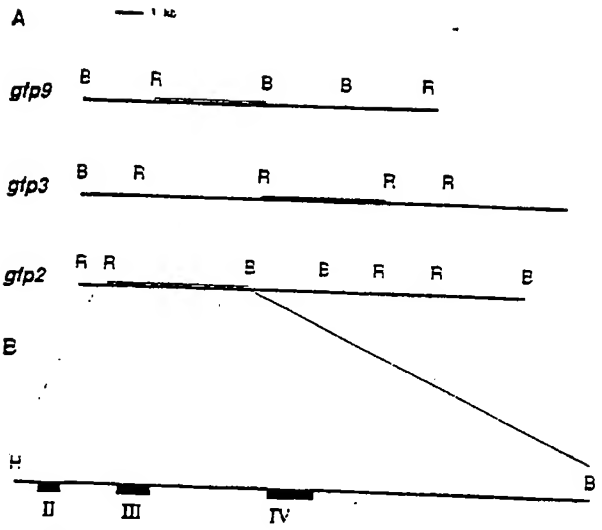

Fig. 4. Restriction enzyme maps of three *Aequorea gfp* genes. (A) The maps of three representative genomic clones are compared. The double lines represent those DNA fragments which hybridize to *gfp1* cDNA. Southern-blot analysis indicated three other genomic clones. λGFP1, 4 and 8 (not shown) lack the 3′ end of the gene. (B) The exon/intron arrangement of the gene encoded by λGFP2 was determined by comparing the nt sequences of the 5-kb *Eco*RI-*Bam*HI and the overlapping 1.8-kb *Hin*dIII fragments of λGFP2 and the *Eco*RI insert of λGFP10 cDNA. The exons are represented by the blackened boxes. I, II and III. The GenBank accession No. for the *gfp2* sequence is M62653.

purified from the genomic DNA library. Based on restriction enzyme and Southern-blot analyses, they represent six different isolates having at least three different restriction maps (Fig. 4). When DNA fragments from the 5′- and 3′-ends of the *gfp1* cDNA were used as hybridization probes, all of the genomic clones were found likely to contain the 5′-end of the gene, but only *gfp2*, *3*, and *9* also contained the 3′ end. The three types of genomic clones are consistent with the presence of multiple GFP isoforms isolated from *A. victoria* (A. Roth, M. Cutler and W.W.W., unpublished). Since the *A. victoria* genomic DNA used for the genomic library was isolated from a large number of

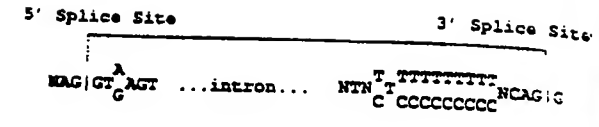

Fig. 5. Alignment of the nt sequences in *gfp2* at the splice junctions. The intron sequences were identified by comparing the nt sequences of *gfp2* and the *gfp10* cDNA (Fig. 2). The consensus sequence is taken from Senapathy et al. (1990).

jellyfish (collected at Friday Harbor, Washington), the three *gfp* genes are representative of the *Aequorea* population as opposed to individual jellyfish.

The *Eco*RI-*Bam*HI and an overlapping *Hin*dIII fragments in the genomic clone λGFP2 (Fig. 4) were sequenced and compared to that of the *gfp10* cDNA to examine the structure of the gene. The *gfp* gene encoded by λGFP2 contains at least three exons spread over 2.6 kb of DNA (Fig. 4). These exons, designated II, III, and IV, encode 64, 98, and 71 aa, respectively. Presumably, a fourth exon is located upstream from the genome since the 15 nt at the 5′ end of the *gfp10* cDNA sequence cannot be aligned to the 5′ region of the DNA sequence derived from the *gfp2* gene. The positions of the introns with respect to the cDNA sequence are indicated (Fig. 2). The aa residues involved in the chromophore are encoded at the 3′ end of exon II. The nt sequences of the *gfp* mRNA splice junctions agree reasonably well with consensus sequences (Fig. 5).

The *gfp10* cDNA is not encoded by the *gfp2* gene since there are several nt differences between their sequences. The nt differences within the protein-coding regions are summarized in Table IA. Four of the 12 single nt differences result in conservative aa replacements at positions 100, 108, 141 and 219 (Table IB). The aa residues encoded at these four positions are consistent with the aa sequences

TABLE I

Sequence differences in the coding regions of the *gfp* clones

A	Nucleotide differences with respect to the *gfp2* gene[a]		B	Amino acid differences[b]			
				aa position[c]	*gfp2* gene	*gfp10* cDNA	*gfp1* cDNA
	gfp10 cDNA	12 (8 silent)					
	gfp1 cDNA	2 (2 silent)					
				100	Tyr	Phe	Tyr
				108	Ser	Thr	Ser
				141	Met	Leu	Met
				219	Ile	Val	

[a] Total number observed upon comparison of the nt sequences of the ORFs in the *gfp* cDNAs with the homologous sequences in the *gfp2* gene.
[b] Observed upon comparison of the translations of the ORFs of both cDNAs and the exons of the *gfp2* gene. The aa numbering is the same as that used in Fig. 2.

bserved in GFP-derived peptides which showed a Tyr at position 100, a Met at position 141, but a Thr at position 08. Eight additional nt differences occur with the *gfp2* gene n the 3'-non-translated region of the *gfp10* cDNA (data o. shown). It is not known whether the *gfp10* cDNA epresents an allele of *gfp2* or another *gfp* gene.

These results will enable us to construct an expression ector for the preparation of non-fluorescent apoGFP. Since no information is yet available regarding the biosynhesis of the chromophore, a recombinant form of this rotein will be a valuable reagent with which to examine the iochemistry of chromophore formation in this unique class f proteins and the mechanism of energy transfer between equorin and GFP.

ACKNOWLEDGEMENTS

We want to extend our thanks to Bonnie Woodward, Darlene Bianca, and Richard McCann for their excellent echnical assistance. Supported in part by a Mellon Award rom the Woods Hole Oceanographic Institution (27/50.44) nd a grant from the American Cancer Society (NP640) to D.C.P. A paper of the journal series New Jersey Agricultural Experiment Station. This work was performed as part of NJAES Project No. 01102. A special thanks goes to Dr. A.O.D. Willows, Director, Friday Harbor Laboratories, for use of laboratory facilities.

REFERENCES

Bokman, S.H. and Ward, W.W.: Renaturation of *Aequorea* green-fluorescent protein. Biochem. Biophys. Res. Commun. 101 (1981) 1372–1380.

Dale, R.M.K., McClure, B.A. and Houchins, J.P.: A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S rDNA. Plasmid 13 (1985) 31–40.

Gubler, U. and Hoffman, B.J.: A simple and very efficient method for generating cDNA libraries. Gene 25 (1983) 263–269.

Huynh, T.V., Young, R.A. and Davis, R.W.: Constructing and screening cDNA libraries in λgt10 and λgt11. In: Glover, D.M. (Ed.), DNA Cloning: A Practical Approach, Vol. 1. IRL Press, Oxford, 1985, pp. 49–78.

Karn, J., Matthes, H.W.D., Gait, M.J. and Brenner, S.: A new selective phage cloning vector, λ2001, with sites for *Xba*I, *Bam*HI, *Hin*dIII, *Eco*RI, *Sst*I and *Xho*I. Gene 32 (1984) 217–224.

Maniatis, T., Fritsch, E.F. and Sambrook, J.: Molecular Cloning, A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.

Morise, J.G., Shimomura, O., Johnson, F.H. and Winant, J.: Intermolecular energy transfer in the bioluminescent system of *Aequorea*. Biochemistry 13 (1974) 2656–2662.

Prasher, D., McCann, R.O. and Cormier, M.J.: Cloning and expression of the cDNA coding for aequorin, a bioluminescent calcium-activated protein. Biochem. Biophys. Res. Commun. 126 (1985) 1259–1268.

Prendergast, F.G. and Mann, K.G.: Chemical and physical properties of aequorin and the green-fluorescent protein isolated from *Aequorea forskalea*. Biochemistry 17 (1978) 3448–3453.

Sanger, F., Nicklen, S. and Coulson, A.R.: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74 (1977) 5463–5467.

Senapathy, P., Shapiro, M.B. and Harris, N.L.: Splice junctions, branch point site, and exons: sequence statistics, identification, and applications to genome project. Methods Enzymol. 183 (1990) 252–278.

Shimomura, O.: Structure of the chromophore of *Aequorea* green fluorescent protein. FEBS Lett. 104 (1979) 220–222.

Thomas, P.S.: Hybridization of denatured RNA transferred or dotted to nitrocellulose paper. Methods Enzymol. 100B (1983) 255–266.

Ward, W.W.: Energy transfer processes in bioluminescence. Photochem. Photobiol. Rev. 4 (1979) 1–57.

Ward, W.W. and Bokman, S.H.: Reversible denaturation of *Aequorea* green-fluorescent protein: physical separation and characterization of the renatured protein. Biochemistry 21 (1982) 4535–4550.

Ward, W.W. and Cormier, M.J.: In vitro energy transfer in *Renilla* bioluminescence. J. Phys. Chem. 80 (1976) 2289–2291.

Ward, W.W. and Cormier, M.J.: An energy transfer protein in coelenterate bioluminescence. J. Biol. Chem. 254 (1979) 781–788.

Ward, W.W., Cody, C.W., Hart, R.C. and Cormier, M.J.: Spectrophotometric identity of the energy-transfer chromophores in *Renilla* and *Aequorea* green-fluorescent proteins. Photochem. Photobiol. 31 (1980) 611–615.

Ward, W.W., Cody, C.W., Prasher, D.C. and Prendergast, F.G.: Sequence of the chemical structure of the hexapeptide chromophore of *Aequorea* green-fluorescent protein. Photochem. Photobiol. 49 (1989) 62S.

Wood, W.I., Gitschier, J., Lasky, L.A. and Lawn, R.M.: Base composition-independent hybridization in tetramethylammonium chloride: a method for oligonucleotide screening of highly complex gene libraries. Proc. Natl. Acad. Sci. USA 82 (1985) 1585–1588.

Yanisch-Perron, C., Vieira, J. and Messing, J.: Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33 (1985) 103–119.